



Research Article

ISSN: 0975-7384
CODEN(USA): JCPRC5

QSAR Modeling and Molecular Docking Studiess on 2-Aminothiazole Derivatives as Potential Inhibitors of Lipase B for the Treatment of *Mycobacterium tuberculosis*

Yakubu Ya'u Muhammad^{1*} Adamu Uzairu², Balarabe Sagagi¹

¹Department of Chemistry Kano University of Science and Technology, Kano, Nigeria

²Department of Chemistry, Ahmadu Bello University Zaria, Zaria, Niger

Received: 13-Jan-2020, Manuscript No. JOCPR-20-6317; **Editor assigned:** 16-Jan-2020, PreQC No. JOCPR-20-6317 (PQ); **Reviewed:** 30-Jan-2020, QC No. JOCPR-20-6317; **Revised:** 02-Jan-2023, ManuscriptNo. JOCPR-20-6317 (R); **Published:** 30-Jan-2023, DOI: 10.37532/0975-7384.2023.15(1).46.

ABSTRACT

Tuberculosis caused by *Mycobacterium tuberculosis* remains a pivotal cause of high mortality worldwide. Inhibitors of *Mycobacterium tuberculosis* lipase B hold promises for treating tuberculosis disease. Computer-aided drug design approaches have proven to be effective in speeding up the process of discovering potential therapeutic agents. Here, a combination of Quantitative Structure-Activity Relationship (QSAR) and molecular docking studies were carried out to investigate anti-tuberculosis potential of 2-Aminothiazole derivatives. A total of 39 2-Aminothiazole derivatives were optimized using Density Function Theory (DFT) method with B3LYP via 6-31G* basis set. Using Genetic Function Algorithm (GFA) available in Material Studio a total of 5 QSAR models were developed. The best model (model 1) was found to have the correlation coefficient, $R^2=0.9452$, $R_{adj}=0.9322$, the model strength, $Q^2=0.9203$. The external validation test used for confirming the predictive power of the built model has $R^2_{pred}=0.6864$, confirming that the model was robust, potentially highly predictive, and satisfactory. The molecular docking study between the ligand and the lipase B receptor (5X7K) showed that compounds 31 and 39 bound to the enzyme with low binding energy values of -8.5 and -8.2 kcal/mol, respectively. Taken together, the QSAR modelling and molecular docking results suggest that these scaffolds may be may be used as lead compounds for the design of new anti-tuberculosis agents.

Keywords: QSAR modeling; Molecular docking; Lipase B; 2-Aminothiazole

INTRODUCTION

Tuberculosis remains the deadliest infectious disease in the world, caused by mycobacterium tuberculosis which the bacteria responsible for the death of 1.5 million people every year [1]. Currently one-third of the world harbour the latent form of *Mycobacterium tuberculosis*, with a lifelong risk of activation and disease development, particularly in people co-infected with HIV [2]. It mainly affects the lungs and other parts of the body such as spine, kidney, and brain on prolong exposure [3]. There current drugs used for the treatment of tuberculosis include isoniazid, rifampicin, ciprofloxacin, pyrazinamide and ethambutol. The tedious duration of therapy and resistance developed by the pathogens are associated with recurrence of the disease, especially as MDR-TB and XDR-TB which pose a global challenge in tuberculosis chemotherapy [4].

It is believed that recent advances in computational chemistry could lead to development of a new drug [5]. Computational methods which reduce the cost for effective evaluation of large virtual data base of chemical compounds are currently employed in designing drug candidates. Some of such approaches include Quantitative Structure-Activity Relationships (QSAR) modeling, Artificial Neural Networks (ANN) analysis, complex networks theory and Machine Learning (ML) [6].

Quantitative structural activity relationship is a statistical model of correlation between a molecular descriptors and experimental activity of a given compound, these descriptors can be in form of either two or three dimension (2D or 3D) [7]. The key success of the QSAR method is the possibility to predict the properties of new chemical compounds prior to synthesis and biological activity testing. This technique is broadly utilized for the prediction of physicochemical properties in the chemical, industrial, pharmaceutical, biological, and environmental spheres [8]. Complimentarily, molecular docking is a computational method used to predict the binding mode of a ligand to a given receptor [9].

Molecular docking studies have been applied to predict the binding affinities of different compounds and to clearly specify the areas of interaction between the ligands and the receptor [10]. QSAR and molecular docking together will give information that can be used in developing potential drug candidate [11]. Here, a potentially highly predictive QSAR model of 2-aminothiazole as potent anti-mycobacterium tuberculosis agents was developed. Compounds 31 and 39 were found to have the lowest interaction energy and may therefore serve as scaffolds for the design of new potent inhibitors of mycobacterium tuberculosis lipase B [12].

MATERIALS AND METHODS

Datasets

A total of thirty-nine (39) derivatives of 2-aminothiazole identified from the literature with the inhibitory activity (IC_{50}) expressed in μM were retrieved. These compounds were converted to pIC_{50} [$= -\log IC_{50} \times 10^6$] to reduce data dissemination and enhanced the linearity in the activity values. The structure of the compounds and their activities are shown in Table S1 [13].

Geometry optimization

The two-dimensional structure of the compounds were drawn using chemdraw 3D Pro 12.0.1V, software and the Spartan 14 v1.1.14 software was used to determine the conformation of the compounds. The structure of the compounds were first minimized by Molecular Mechanic Force Field (MMFF) to eliminated the strain energy, then the optimization was performed using Density Functional Theory (DFT) with B3LYP and 6-31G* basis set in Spartan software [14].

Descriptors calculation

The thirty nine optimized molecules were saved in sdf form and the descriptors of the optimized compounds were calculated by using the descriptor toolkit of PaDEL software version 2.20 [15].

Data normalization and pre-treatment

The values of the descriptors were normalized so as to give all the variables the equal chance of influencing the model. Noise and redundant data were removed by subjecting the data to pre-treatment [16].

$$X = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

Where X_i is the value of each descriptor for a given molecule, $X_{\max}$ is the maximum value for all the column of the descriptor X, while $X_{\min}$ is the minimum value for each column of descriptor X [17].

Training and test set

The dataset was split into training set and test set with the aid of Kennard and Stone's algorithm. The training set comprises 27 compounds and was used to build the model while the remaining 12 compounds serve as a test set used to validate the built model [18].

Internal validation

Internal validation of the model was carried out using Materials Studio version 8 software, employing the Genetic Function Approximation (GFA) method. The models were estimated using the LOF. The LOF is measured using a slight variation of the original Friedman formula, so that the best fitness score could be received. LOF is expressed as follows:

$$LOF = \frac{SEE}{(1 - (C + d \times p) / M)} \quad (2)$$

Where SEE means the standard error of estimation, C is defined as the number of terms in the model, d is a user smoothing parameter, p is the number of descriptors that appear in the model, and M is the amount of data in the training set. SEE is a measure of model's quality-the lower the value of SEE the better the quality of the model [19]. SEE is defined as

$$SEE = \sqrt{R^2} \quad (3)$$

The square of the correlation coefficient (R^2) measure the power of the model which explain how the activity values of the compounds used in building the model vary. A satisfied model has an R^2 value of 1, and the more the value of the R^2 deviate from 1, the more the robustness of the model reduces that is the closer is the value of R^2 to 1 the better the developed model [20].

$$R = 1 - \left[\frac{\sum(Y_{exp} - Y_{pred})^2}{\sum(Y_{exp} - Y_{train})^2} \right] \quad (4)$$

Where Y_{exp} , Y_{pre} and $T_{training}$ are the experimental, predicted and the mean experimental activities of the samples in the training set respectively. R^2 values varies with increase in the number of descriptors, and this makes R^2 unreliable in measuring the fitness of the model. Thus R^2 is adjusted for all the number of variables in the model and it is defined as:

$$R^2_{adj} = \frac{R^2 - k(n-1)}{n-p+1} \quad (5)$$

Where k is the number of independent variables in the model and n represents the number of descriptors. The strength of the equation of QSAR to predict activity of a compound was assessed using leave-one-out cross-validation method with the revised formula below:

$$Q^2 = 1 - \left[\frac{\sum(Y_{exp} - Y_{pred})^2}{\sum(Y_{exp} - Y_{train})^2} \right] \quad (6)$$

External validation

Internal validation of a model is employed to evaluate the predictive ability and stability of the model, however, no real predictive capacity is shown for the external samples. This necessitates the need to ascertain the predictive ability externally and as well as extrapolation. The predictive R^2 (R^2 test) is calculated as follows:

$$R^2_{test} = 1 - \sum Y_{pred}(\text{test}) \text{ and } Y_{exp}(\text{test}) \quad (7)$$

Where $Y_{pred}(\text{test})$ and $Y_{exp}(\text{test})$ are the predicted and experimental activity test sets respectively [21].

Applicability domain

Applicability domain of a QSAR model is employed to determine outliers and influential compounds and to affirm the reliability and robustness of the model generated. The leverage is one of the most important techniques used to evaluate the applicability domain of the QSAR model, and for a given chemical compound is defined as follows:

$$h_i = x_i \quad (i=k, \dots, p) \quad (8)$$

Where x_i is the training compound matrix I , X is $n \times k$ descriptor matrix of the training set compounds and X^T is the transpose matrix X used to build the model. As a prediction tool, the warning leverage (h^*) which is the limit for X values is defined as:

$$h^* = \frac{3(p+1)}{n} \quad (9)$$

Where n is the number of training compounds, and p is the number of descriptors in the model.

Molecular docking studies

The optimized compounds were converted into PDB format using spartan software. The structure of the enzyme (Lipase B) with the Protein Data Bank (PDB) code 5X7k was retrieved. Discovery studio software was used to prepare the protein based on the protonation state of the titratable residues; to delete water molecules and ions and to minimize the energy of the structure.

All the compounds were docked into the active site lipase B enzyme using Autodock vina 4.2 (PyRx) virtual screening software. The grid box centre and dimension were chosen automatically by the program.

RESULTS AND DISCUSSION

QSAR results

The best model equation is given below:

Model 1:

$$pIC50 = -0.087483399 * VE3_{Dzp} - 1.769976361 * SpMax7_{Bhv} - 0.679599256$$

$$R^2 = 0.9452, R^2_{adj} = 0.9322, Q^2_{loo} = 0.9203, N_{train} = 27, R^2_{test} = 0.6864, N_{test} = 12$$

The higher values of R^2 (0.9452), R^2_{adj} (0.9322) and Q^2 (0.9203) of the model indicate the good assessment of the internal validation of the model. R^2 for the assessment of external validation of the model for the test set was found to be 0.6864. Table S2 give the symbols, description and classes of the descriptors used in the model 1. Table 1 shows the external validation and calculation of the predictive R^2 of model 1 (Table 1).

Table 1: External validation of model 1.

VE3_Dzp	SpMax7_Bhv	SpMax8_Bhv	CrippenMR	SpMin3_Bhp	pIC ₅₀	Y _{prd}	Y _{obs} -Y _{pred}	(Y _{obs} -Y _{pred}) ²	Y _{train}	Y _{obs} -Y _{train}	(Y _{obs} -Y _{train}) ²
-3.70467	2.496073	2.153689	79.2887	1.549596	5.67	5.918751	24875	0.061877	5.32	0.35	0.1225
-2.95077	2.476824	2.096006	76.2737	1.568992	6.09	6.053353	0.036647	0.001343	5.32	0.77	0.5929
-3.59436	2.493238	2.134074	78.2787	1.469239	5.43	5.012456	0.417544	0.174343	5.32	0.11	0.0121
-8.85959	2.477437	2.47895	96.9787	1.461846	5.61	5.722049	-0.11205	0.012555	5.32	0.29	0.0841
-5.30468	2.464133	2.483333	95.7847	1.497212	5.53	5.78728	-0.2573	0.066193	5.32	0.21	0.0441
-2.55742	2.405455	2.405455	76.3957	1.546924	5.55	5.69514	-0.14515	0.02107	5.32	0.23	0.0529
-9.71179	2.445815	2.504837	101.4062	1.526405	6.15	6.6739	-0.5239	0.274478	5.32	0.83	0.6889
-1.21909	2.291151	2.040559	73.5737	1.550146	5.85	5.9829	-0.1329	0.017654	5.32	0.53	0.2809
-6.28138	2.412242	2.371706	85.9922	1.543622	5.92	6.271	-0.351	0.123183	5.32	0.6	0.36
-6.91556	2.454399	2.478702	99.7622	1.508849	5.85	6.1911	-0.3411	0.116337	5.32	0.53	0.2809
-4.15689	2.446377	2.47365	88.1165	1.545558	5.79	6.0378	-0.2478	0.061382	5.32	0.47	0.2209
-3.40024	2.634041	2.359466	83.9327	1.556345	5.92	5.7158	0.20418	0.04169	5.32	0.6	0.36
								$\sum(Y_{obs}-Y_{pred})^2 = 3.1002$			$\sum(Y_{obs}-Y_{train})^2 = 3.1002$

The plot of predicted activities of both training and test sets against experimental activities are shown in (Figure 1). The high linearity of the plot indicates the reliability of the model 1 which suggests its high predictive power.

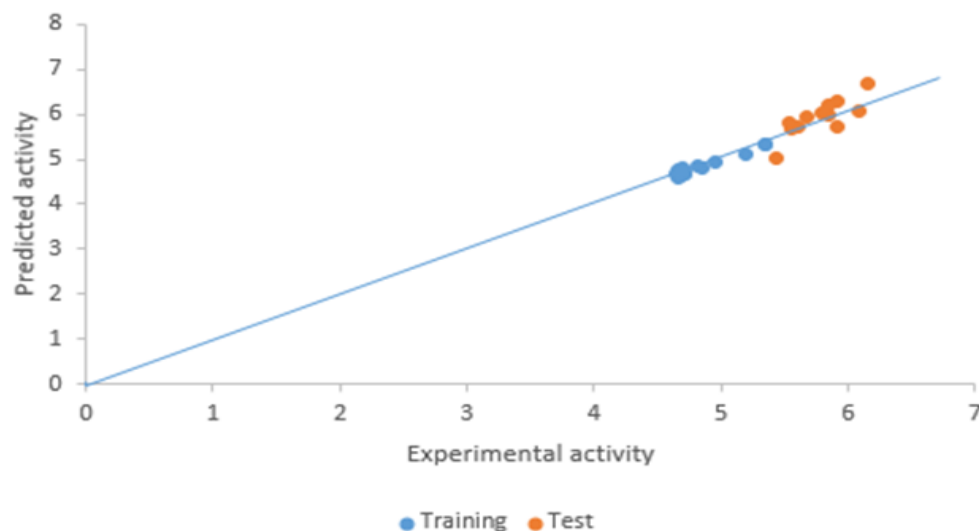


Figure 1: The plot of the experimental and predicted activity of both the training and test sets for model 1.

Similarly, a plot of experimental activity values against the standardized residuals was shown in Figure 2. Interestingly, the symmetric and random propagation of the standardized residuals of data points are on both side of zero and this means that there was no systematic error in the built model (Figure 2).

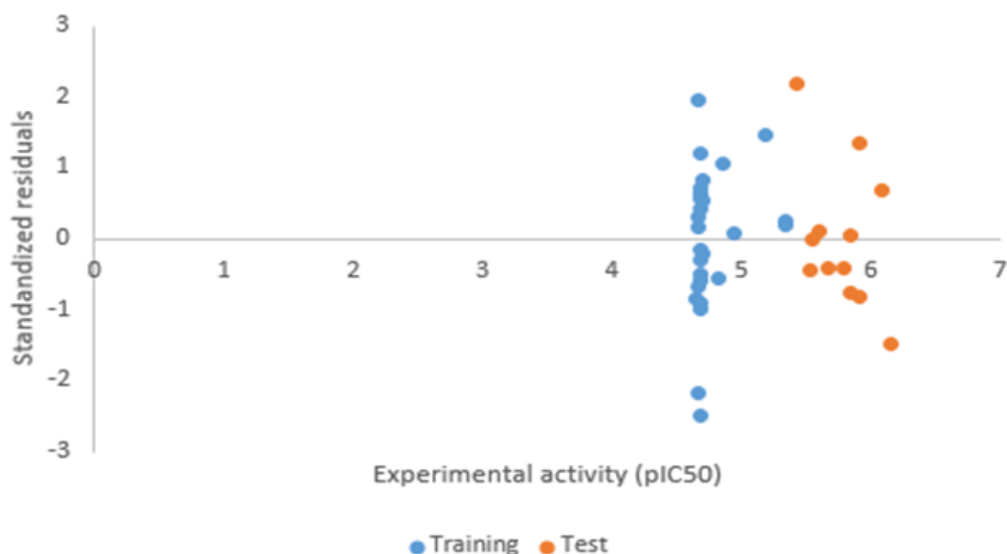


Figure 2: Plot of standardized residual versus experimental activity.

Furthermore, Table S3 shows the experimental and predicted activities with their residual values. The lower residual value between experimental and the predicted activities confirm the predictive ability and capacity of the model. The Williams plot of the standardized residuals versus leverages is presented in Figure 3. From the results, it is evident that one compound from the training set has the leverages high than the warning leverages ($h^*=0.66$) and that is considered to be outlier compound, due to its different structure from the other compounds of the dataset.

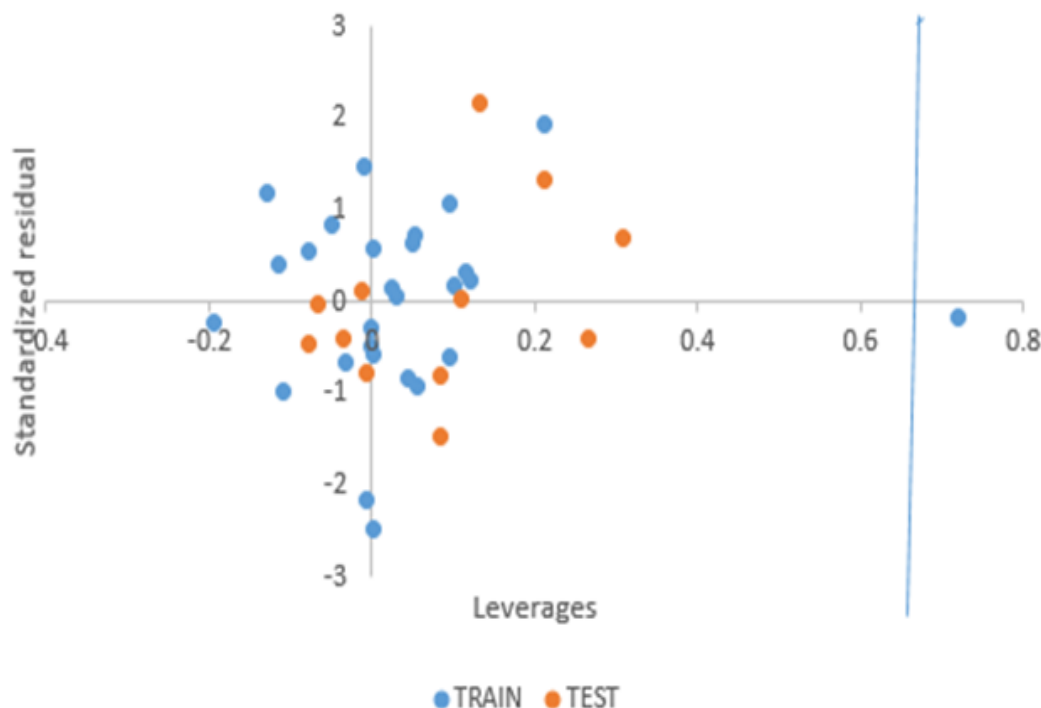


Figure 3: Williams plot of standardized residual and leverages of both training and test sets of the model.

Molecular docking results

Molecular docking studies were carried out between the lipase B and the 39 compounds of 2-Aminothiazoles derivatives to predict their binding mode and binding energy scores. The binding energy values from -5.4 to -8.5 kcal/mol, and the compounds 10, 26, 30, 31 and 39 were predicted to have the highest docking scores in ascending order. Compound 39 showed the lowest binding energy, -8.5 kcal/mol and formed hydrogen interaction with ARG480 (2.82796 Å), GLY405 (2.51971 Å) and GLY405 (3.21025 Å) active site. In addition, other interaction with active site residues residues, LYS401, LYS401, TRY410, PRO412, VAL379, LEU404, ILE408, PRO412, and PRO412 were observed (Table 2 and Figure 4).

Table 3: Binding energy, interaction residues, hydrogen bonds and hydrogen bond distance of ligand with highest docking scores.

Ligands	Binding energy (kcal/mol)	Interaction residues	Hydrogen bond	Hydrogen bond distance (Å)
10	-8	ILE381,PRO412,TYR410,PRO412,ARG428,VAL379,ILE408	TYR410	2.8801
26	-8	ARG480, PHE418, LYS401, VAL379, LEU404, ILE408	ARG480,ARG428, LYS401,LYS401, ARG428	2.68854,2.0322,2.27732,2.29581,2.35388
30	-8	TYR410, ARG480, ASP402, LYS401, VAL379, ILE408	ARG428, ARG428 and ARG428	2.32392,2.61343,2.63796
31	-8.5	LYS401, LYS401, TYR410, PRO412, VAL379, LEU404, ILE408, PRO412, and PRO412	ARG480,GLY405, and GLY405	2.82796,2.51971,3.21025
39	-8.2	ARG480, ARG428, LYS401, VAL379, EU404, and ILE408	ARG428,LYS401, and ARG428	2.627,2.38907,and 2.37479

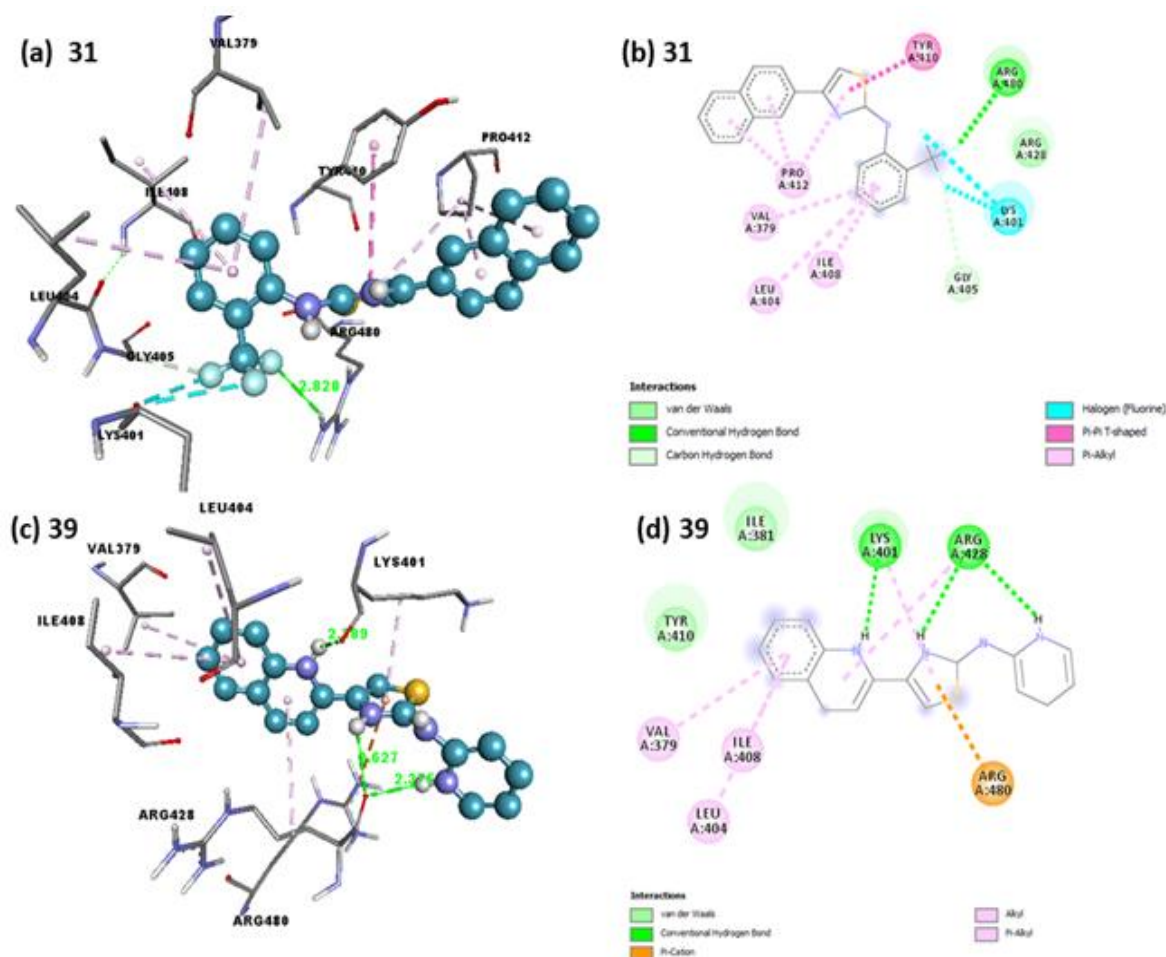


Figure 4: a) 3D and b) 2D interaction diagrams of predicted complexes of lipase B with compound 31 (binding energy=-8.5 kcal/mol); c) 3D and d) 2D interaction diagrams of predicted complexes of lipase B with compound 39 (binding energy=-8.2 kcal/mol).

DISCUSSION

QSAR was performed to investigate the structure activity relationship of the inhibitory compounds as potent anti-mycobacterium tuberculosis. In this study the R^2 value recorded in the predicted activity against experimental activity of both training and test set shown in Figure 1 were in agreement with GFA derived R^2 value reported by other [25]. The plot of standardized residual verses experimental activity shown in Figure 2 indicates that there was no systematic error in the model built as the spread of standardized residual values were on both sides of zero [26] and the Williams plot of the standardized residuals shown in figure.3 were in agreement with the finding of other [27] with the applicability domain of the square area of ± 3 and the warning leverage of ($h^*=0.66$). The molecular docking result in this research has shown that the binding score was in agreement with finding of other researcher which were better than the commercially sold anti-mycobacterium tuberculosis; isoniazid (-5.3 kcal/mol) and ethambutol (-5.8 kcal/mol).

CONCLUSION

QSAR and molecular docking studies of a total of 39 compounds of 2-aminothiazoles derivatives as lipase B inhibitors were performed. The Genetic Function Algorithm (GFA) to build a total of five models. The best model was found to have $R^2=0.9452$, $R_{adj}=0.9322$, $Q^2=0.9203$ and the external validation $R^2_{pred}=0.6864$. From the molecular docking studies carried out shows that all the compounds binding with target receptor favourably. Ligand

31 has the highest binding energy of -8.5 kcal/mol showed hydrogen interaction with ARG480 (2.82796 Å), GLY405 (2.51971 Å) and GLY405 (3.21025 Å) active site and form interaction residues with active site of the receptor LYS401, LYS401, TRY410, PRO412, VAL379, LEU404, ILE408, PRO412, and PRO412. The QSAR model generated provides a valuable approach for ligand base design, while molecular docking studies give the valuable for structure base design. These two approaches will significantly help pharmaceutical and medicinal chemist to design new anti-Mykobacterium tuberculosis agent.

ACKNOWLEDGEMENT

The authors wish to acknowledge the management of Kano State University of Science and Technology Wudil, and the member of the department of chemistry, Ahmadu Bello University Zaria. My sincerely acknowledge to Ibrahim Ahmad Tijjani and Sagir Yusif Ismail for their technical support and advice.

REFERENCES

1. Anil K, Eric A, Nacer L, et al. *Nature*. **2011**;469.
2. Cramer RD, Patterson DE, Bunce JD. *J Am Chem Soc*. **1988**;110:599-596.
3. Speck-Planche A, Tullius SM, Paulo-Emerenciano V. *Curr Pharm Des*. **2010**;16:2656-2665.
4. Shola EA, Sani U, Uzairu A. *J King Saud Univ Sci*. **2018**;1-8.
5. Das J, Chen P, Norris D, et al. *J Med Chem*. **2006**;49(23):6819-6832.
6. Uba AI, Yelekçi K. *Comput Biol Chem*. **2018**;75:131-142.
7. Boukarai Y, Khalil F, Bouachrine M. *Diab Compl*. **2017**.
8. Muhammad TI, Adamu U, Gideon AS, et al. *J King Saud Univ Sci*. **2018**;1-10.
9. Jain S, Ghate M, Bhadoriya K, et al. *Organ Med Chem Lett*. **2012**;2:1.
10. Kesicki EA, Bailey MA, Ovechkina Y, et al. *Plos One*. **2016**;11(5):e0155209.
11. Abdulfatai U, Uzairu A, Uba, S. *Cogent Chem*. **2016**;2.
12. Yap CW. *J Comput Chem*. **2011**;32:1466-1474.
13. Kennard RW, Stone LA. *Technomet*. **1969**;11(1):137-148.
14. Friedman JH. *Annals Stat*. **1991**;19(1):1-67.
15. Adedirin O, Uzairu, Shallangwa GA, et al. *J Eng Exact Sci JCEC*. **2018**;4(2).
16. Bello AS, Uzairu A, Shallangwa GA, et al. *J Eng Exact Sci JCEC*. **2018**;4(2).
17. Tropsha A, Gramatica P, Gombar VK. *Mol Inform*. **2003**;22:69-77.
18. Trott O, Olson AJ. *J Comput Chem*. **2010**;31:455-461.
19. Jalali-Heravi M, Kyani A. *J Chem Info Comp Sci*. **2011**;44(4):1328-1335.
20. Shola EA, Sani U, Uzairu A. *Hindawi J Pathogens*. **2018**;1018694.
21. Li Z, Wan H, Shi Y, et al. *J Chem Inf Comput Sci*. **2004**;44:1886-1890.